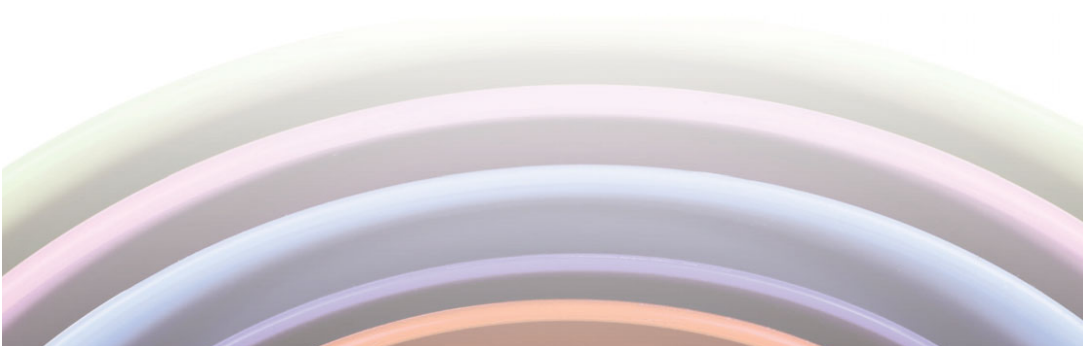


14

Secondary analysis and official statistics

Chapter outline

Introduction	311
Other researchers' data	312
Advantages of secondary analysis	312
Limitations of secondary analysis	315
Accessing the Data Archive	316
Official statistics	320
Reliability and validity	322
Condemning and resurrecting official statistics	324
Official statistics as a form of unobtrusive method	325
Key points	327
Questions for review	327





Chapter guide

This chapter explores the possibilities associated with the analysis of data that have been collected by others. There are two main types discussed in this chapter:

- the secondary analysis of data collected by other researchers;
- the secondary analysis of official statistics—that is, statistics collected by government departments in the course of their work or specifically for statistical purposes.

This chapter explores:

- the advantages and disadvantages of carrying out secondary analysis of data collected by other researchers, particularly in view of many data sets being based on large, high-quality investigations that are invariably beyond the means of students;
- how to obtain such data sets;
- the potential of official statistics in terms of their reliability and validity;
- the growing recognition in recent years of the potential of official statistics after a period of neglect as a result of criticisms levelled at them;
- the notion that official statistics are a form of *unobtrusive method*—that is, a method that is not prone to a reaction on the part of those being studied to the fact that they are research participants.

Introduction

Many of the techniques we have covered so far—survey research by questionnaire or structured interview, structured observation, and content analysis—can be extremely time-consuming and expensive to conduct. Students in particular may have neither the time nor the financial resources to conduct very extensive research. Yet we know that large amounts of quantitative data are collected by social scientists and others. Moreover, many organizations, most notably government departments and their various representatives, collect data that are presented in statistical form and that may be usable by social scientists. Would it not be a good idea to analyse such data rather than collect new data? It would have the additional advantage for the long-suffering public that they would not be bothered by interviewers and by questionnaires popping through their letter boxes.

This is where *secondary analysis* comes in. Secondary analysis offers this kind of opportunity. Key concept 14.1 contains a brief definition of secondary analysis and raises one or two basic points about what it involves. As the opening paragraph suggests, we will in this chapter be concerned with two kinds of issue:

1. the secondary analysis of data that have been collected by other researchers;
2. the secondary analysis of data that have been collected by various institutions in the course of their business.

One prominent form of such data is official statistics concerned with areas of social life such as crime, unemployment, and strikes. The use of such official statistics for social research has been controversial, and aspects of the ensuing debate will be addressed below.



Key concept 14.1

What is secondary analysis?

Secondary analysis is the analysis of data by researchers who will probably not have been involved in the collection of those data, for purposes that in all likelihood were not envisaged by those responsible for the data collection. Secondary analysis may entail the analysis of either quantitative data (Dale et al. 1988) or qualitative data (Corti et al. 1995), but it is with the former that we will be concerned in this chapter. To some extent, it is difficult to know where primary and secondary analysis start and finish. If a researcher is involved in the collection of survey interview data and analyses some of the data, resulting in some publications, but then some time later decides to rework the data, it is not entirely clear how far the latter is primary or secondary analysis. Typically, secondary analysis entails the analysis of data that others have collected, but, as this simple scenario suggests, this need not necessarily be the case.

For example, I was involved in a secondary analysis of data on dietary choices, attitudes, and practices among a sample of Leicestershire residents (Beardsworth et al. 2002). This analysis reflected my and the lead author's research interests. The analysis dealt with the significance of gender in nutritional attitudes and choices, which had not been the focus of previous analyses of these data. I had not been involved in the collection of the data nor in the design of the research, but the co-authors of the article that reported the analysis had been involved in these phases of the research. It is questionable whether for them it was a secondary analysis, but for me it certainly was. The analysis shows that women are more likely than men to adopt what we call a 'virtuous' pattern of eating, which acknowledges a sensitivity to, and preparedness to act in terms of, ethical and nutritional principles that are respected in many Western societies. The article also shows that these principles have a sinister component in that they frequently come to be associated among women with feelings of guilt, disordered eating patterns, and concerns about body shape.



Other researchers' data

There are several reasons why secondary analysis should be considered a serious alternative to collecting new data. These advantages of secondary analysis have been covered by Dale et al. (1988), from which I have borrowed most of the following observations. In considering the various advantages of secondary analysis, I have in mind the particular needs of the lone student conducting a small research project as an undergraduate or a more substantial piece of work as a postgraduate. However, this emphasis should definitely not be taken to imply that secondary analysis is really appropriate or relevant only to students. Quite the contrary: secondary analysis should be considered by all social researchers, and, indeed, the Economic and Social Research Council (ESRC) requires applicants for research grants who are proposing to collect new data to demonstrate that relevant data are not already available in the UK Data Archive (see below).

My reason for emphasizing the prospects of secondary analysis for students is simply based on my personal experience that they tend to assume that any research they carry out has to entail the collection of primary data. Unless you are advised otherwise by your supervisor or

by any documentation supplied by your department or university, it is worth giving serious consideration to doing a secondary analysis, because it will allow you to spend more time on the analysis and interpretation of data. Moreover, you simply will not be in a position to produce a data set of comparable quality, because of the lack of time and resources likely to be available to you. Further, as the outline in the next section of the advantages of secondary analysis suggests, you may be able to conduct analyses that would be inconceivable if you relied on data you collected yourself.

Advantages of secondary analysis

Secondary analysis offers numerous benefits to students carrying out a research project. These are outlined below.

- *Cost and time.* As noted at the outset, secondary analysis offers the prospect of having access to good-quality data for a tiny fraction of the resources involved in carrying out a data-collection exercise yourself. Numerous data sets are available from the Data Archive, which is housed at the University of Essex in

Colchester. You do not even have to go to Colchester yourself to search for data. The Archive has a very good website, which can be searched in a variety of ways, such as keywords (see below).

- **High-quality data.** Many of the data sets that are employed most frequently for secondary analysis are of extremely high quality. By this I mean several things. First, the sampling procedures have been rigorous, in most cases resulting in samples that are as close to being representative as one is likely to achieve. While the organizations responsible for these studies suffer the same problems of survey non-response as anybody else, well-established procedures are usually in place for following up non-respondents and thereby keeping this problem to a minimum. Second, the samples are often national samples or at least cover a wide variety of regions of Great Britain or the UK. The degree of geographical spread and the sample size of such data sets are invariably attained only in research that attracts quite substantial resources. It is certainly inconceivable that student projects could even get close to the coverage that such data sets attain. Third, many data sets have been generated by highly experi-

enced researchers, and, in the case of some of the large data sets, like the General Household Survey (GHS), the British Social Attitudes (BSA) survey (see Research in focus 14.1), and the British Household Panel Survey (BHPS; see Research in focus 14.1 and 14.2), the data have been gathered by social research organizations that have developed structures and control procedures to check on the quality of the emerging data.

- **Opportunity for longitudinal analysis.** Partly linked to the last point is the fact that secondary analysis can offer the opportunity for longitudinal research, which, as noted in Chapter 3, is rather rare in the social sciences because of the time and cost involved. Sometimes, as with the BHPS, a panel design has been employed, and it is possible to chart trends and connections over time. Such data are sometimes analysed cross-sectionally, but there are obviously opportunities for longitudinal analysis as well. Also, with data sets such as the GHS and the BSA survey, where similar data are collected over time, usually because certain interview questions are recycled each year, trends (such as shifting opinions or changes in behaviour)



Research in focus 14.1

Religion in Britain and the 'believing without belonging' thesis

'Believing without belonging', as the phrase implies, is meant to suggest that religion has not declined in modern British society in terms of the extent to which religious beliefs are held; rather, the phrase suggests that it is allegiance to institutional religion that has declined, in that religious beliefs are still fairly robust. This thesis was explored by Voas and Crockett (2005), who conducted secondary analyses of both the BHPS and the BSA survey. For the BHPS analysis, the authors took waves 1 (1991) and 9 (1999–2000), focusing on three questions that are asked:

1. A question asking whether the respondent views him- or herself as belonging to any religion and, if so, which one. This is used to indicate a person's religious affiliation.
2. A question asking how often the respondent attends religious services or meetings. This is commonly used as an indicator of participation.
3. A question asking each respondent how much difference religious beliefs make to his or her life. Respondents answer on a scale: no difference; little difference; some difference; great difference. This question is taken to indicate strength of belief and its significance in the respondent's life.

The findings lead Voas and Crockett (2005: 15) to suggest that 'religious decline is principally the result of differences between generations: each age cohort is less religious than the last'. Also, the same data do not support the 'believing without belonging' thesis: if anything, religious belief is declining faster than belonging. They explored the BSA data, which could be examined over a much longer period (1983–2002). The data were examined as a continuous time series for each year for which there were data on religious affiliation and belief. They show that the dominant effect was a cohort effect; that is, each cohort (five cohorts were distinguished) was less religious (in terms of affiliation, attendance, and belief) than the previous one. Overall, then, these longitudinal data suggest that the 'believing without belonging' thesis cannot be supported empirically.



Research in focus 14.2

Household work: an example of secondary analysis

Bond and Sales (2001) conducted a secondary analysis of wave 4 (1994) of the BHPS. They were interested in the organization of household work among couples where both members were of working age and were in paid work. Of the 9,481 adults interviewed during this wave of data collection, 981 couples met the researchers' criteria. Bond and Sales were interested in the variables associated with the amount of time men and women spent in household work and the level of sharing of household jobs. As regards the former of these dependent variables, the authors show, for example, that the amount of time spent in household tasks was negatively associated with the relative economic power of one's partner and that men tended to spend more time on domestic tasks the longer the hours of their partners in paid work. Sharing of tasks was greater among women working longer hours, but men take up only a small amount of the household work that does not get done as a result of women working. Moreover, men with greater economic power than their partners tended to share less than those without such power. A study like this allows an important topic to be illuminated through a relatively representative sample that would require a great deal of effort to compile.

can be identified over time. With such data sets, respondents differ from year to year, so that causal inferences over time cannot be readily established, but nonetheless it is still possible to gauge trends.

- *Subgroup analysis.* When large samples are the source of data (as in the GHS and BHPS), there is the opportunity to study what can often be quite sizeable subgroups. Very often, in order to study specialized categories of individuals, small localized studies are the only feasible way forward because of costs. Large data sets can frequently yield quite large nationally representative samples. For example, Arber and Gilbert (1989) used the 1980 GHS to isolate a sample of over 4,500 elderly people living in private households. In 1980, respondents aged 65 and over were asked various questions about their ability to perform certain tasks. This information was used to compile a disability index. Levels of disability could then be related to patterns of caring for the elderly. As Arber and Gilbert (1989: 75) observe: 'The large sample size, high response rate (82 per cent) and representative nature of the sample size make the GHS a valuable data source to complement, extend and systematically test findings and theoretical ideas derived from small, qualitative and localised studies.' While the data did not address the elderly in institutional care, the survey nonetheless provides a valuable source of high-quality data on the elderly. It is easy to see how a wide range of different subgroups could be identified for similar kinds of analysis.

- *Opportunity for cross-cultural analysis.* Cross-cultural research has considerable appeal at a time when social

scientists are more attuned to the processes associated with globalization and to cultural differences. It is easy to forget that many findings should not be taken to apply to countries other than that in which the research was conducted. However, cross-cultural research presents barriers to the social scientist. There are obvious barriers to do with the cost and practical difficulties of doing research in a different country, especially when language and cultural differences are likely to be significant. The secondary analysis of comparable data from two or more countries provides one possible model for conducting cross-cultural research. In order for a cross-cultural analysis to be conducted, some coordination is necessary so that the questions asked are comparable. The research on religiosity described in Research in focus 2.4 and 7.3 by Kelley and De Graaf (1997) provides an example of such coordination. The authors describe the process as follows:

Data are from the 1991 'Religion' module of the International Social Survey Programme (ISSP), an international consortium composed primarily of academic survey organizations . . . Each year, the ISSP creates a module containing exactly the same questions, answer categories, and sequencing for all countries surveyed. . . . The samples are all large, representative national samples of adults. The most common procedure is to hold face-to-face interviews with a stratified random sample . . . followed by a leave-behind self-completion questionnaire containing the ISSP module . . . (Kelley and De Graaf 1997: 642)

Kelley and De Graaf's results were based upon a secondary analysis of the data from the fifteen nations involved in the research. Opportunities for such cross-cultural analysis appear to be increasing. For example, core questions used in the Labour Force Survey (LFS) are also used in equivalent surveys conducted by EU member states.

- *More time for data analysis.* Precisely because data collection is time-consuming, the analysis of data is often squeezed. It is easy to perceive the data collection as the difficult phase and to take the view that the analysis of data is relatively straightforward. This is not the case. Working out what to make of your data is no easy matter and requires considerable thought and often a preparedness to consider learning about unfamiliar techniques of data analysis. While secondary analysis invariably entails a lot of data management—partly so that you can get to know the data and partly so that you can get it into a form that you need (see below)—and this phase should not be underestimated—the fact that you are freed from having to collect fresh data means that your approach to the analysis of data can be more considered than perhaps it might otherwise have been.
- *Reanalysis may offer new interpretations.* It is easy to take the view that, once a set of data has been analysed, the data have in some sense been drained of further insight. What, in other words, could possibly be gained by going over the same data that someone else has analysed? In fact, data can be analysed in so many different ways that it is very unusual for the range of possible analyses to be exhausted. Several possibilities can be envisaged. A secondary analyst may decide to consider the impact of a certain variable on the relationships between variables of interest. Such a possibility may not have been envisaged by the initial researchers. Second, the arrival of new theoretical ideas may suggest analyses that could not have been conceived of by the original researchers. In other words, the arrival of such new theoretical directions may prompt a reconsideration of the relevance of the data. Third, an alternative method of quantitative data analysis may be employed and offer the prospect of a rather different interpretation of the data. Fourth (and related to the last point), new methods of quantitative data analysis are continuously emerging. Disciplines such as statistics and econometrics are continually developing new techniques of analysis, while some techniques are developed within the social sciences themselves. As awareness of such

techniques spreads, and their potential relevance is recognized, researchers become interested in applying them to new data sets.

- *The wider obligations of the social researcher.* For all types of social research, research participants give up some of their time, usually for no reward. It is not unreasonable that the public should expect that the data that they participate in generating should be mined to its fullest extent. However, much social research is chronically under-analysed. Primary researchers may feel they want to analyse only data relating to central research questions or lose interest as a new set of research questions interpose themselves into their imagination. Making data available for secondary analysis enhances the possibility that fuller use will be made of data.

Limitations of secondary analysis

The foregoing list of benefits of secondary analysis sounds almost too good to be true. In fact, there are not very many limitations, but the following warrant some attention.

- *Lack of familiarity with data.* When you collect your own data, when the data set is generated, it is hardly surprising that you are very familiar with the structure and contours of your data. However, with data collected by others, a period of familiarization is necessary. You have to get to grips with the range of variables, the ways in which the variables have been coded, and various aspects of the organization of the data. The period of familiarization can be quite substantial with large complex data sets and should not be underestimated.
- *Complexity of the data.* Some of the best-known data sets that are employed for secondary analysis, such as the BHPS, are very large in the sense of having large numbers of both respondents and variables. Sometimes, the sheer volume of data can present problems with the management of the information at hand, and, again, a period of acclimatization may be required. Also, some of the most prominent data sets that have been employed for secondary analysis are known as *hierarchical* data sets, such as the GHS and the BHPS. The difficulty here is that the data are collected and presented at the level of both the household and the individual, as well as other levels. The secondary analyst must decide which level of analysis is going to be employed. If the decision is to analyse individual-level data, the individual-level data must then be extracted from the data set. Different data will apply to each

level. Thus, at the household level, the GHS provides data on such variables as number of cars and consumer durables, while, at the individual level, data on income and employment can be found. Dale (1987) was interested in life-cycle stages and employed household-level data from the 1979 GHS to develop a typology of life-cycle stages, which included fourteen categories, and various correlates of the various categories, such as net disposable household income.

- *No control over data quality.* The point has been made on several occasions that secondary analysis offers the opportunity for students and others to examine data of far higher quality than they could collect themselves. However, this point applies mainly to data sets such as the GHS, the BHPS (see Research in focus 3.10), the BCS (see Thinking deeply 7.1), the NCDS (see Research in focus 3.11), the Social Change and Economic Life Initiative (see Research in focus 7.1), and the BSA survey (see Research in focus 2.4). While the quality of data should never be taken for granted, in the case of such data sets it is reasonably assured, though that is not to say that they will necessarily meet all of a prospective secondary analyst's needs, since data may not have been collected on an aspect of a topic that would have been of considerable interest. With other data sets, somewhat more caution may be necessary in connection with data quality, although certain fundamental checks on quality are usually made by archives in which data are deposited.

- *Absence of key variables.* Because secondary analysis entails the analysis of data collected by others for their own purposes, it may be that one or more key variables may not be present. You may, for example, want to examine whether a relationship between two variables holds even when one or more other variables are taken into account. Such an analysis is known as **multivariate analysis**, an area that will be touched on in the next chapter. The inability to examine the significance or otherwise of a theoretically important variable can be frustrating and can arise when, for example, a theoretical approach that has emerged since the collection of the data suggests its importance. Obviously, when researchers collect primary data themselves, the prospect of this happening should be less pronounced.

Accessing the Data Archive

The UK Data Archive at the University of Essex is likely to be your main source of quantitative data for secondary analysis, although it may be that some of your lecturers will have data sets that they are prepared to put at your disposal. The best route for finding out about data held at the Archive is to examine its online catalogue. Access to this catalogue can be obtained by going to the UK Data Archive home page at:

www.data-archive.ac.uk (accessed 28 January 2011),

Plate 14.1

The Economic and Social Data Service Search Catalogue

The screenshot shows the 'Economic and Social Data Service' website. At the top, there is a navigation bar with links: Home | A-Z index | Site map | Contact | Login | Search. Below this is a search bar with the text 'Search site/data'. The main content area is titled 'Search' and contains a search form. The form has several sections: 'Title' with a search box containing 'nat', 'Author' with a search box, 'AND' with a dropdown menu set to 'All Fields', 'AND Year (yyyy)' with a search box, 'AND Service' with a dropdown menu set to 'All Data Catalogue', and 'Sort by' with a dropdown menu set to 'Relevance'. There are also checkboxes for 'All Data Catalogue' and 'Show only major studies'. A 'Clear Search' button is located at the bottom left of the search form. On the right side of the search form, there are radio buttons for 'Any', 'All', and 'Exact' for each search field. Below the search form, there is a link: 'Visit the UK Data Archive for other related catalogues.' On the left side of the page, there is a sidebar with the heading 'Finding data' and a list of links: Overview, Catalogue search, Help on searching, About catalogue, Dictionary of terms, Variables search, Browse by subject, Major studies, Major datasets, New releases, HESSET thesaurus, Other archives, About the data, and Login. Below this sidebar, there is a section titled 'NEW USERS' with a small image and the text: 'How do I find data? How do I register? What data are available? How frequently asked questions...'

or by going directly to the Economic and Social Data Service (ESDS) catalogue search page at:

www.esds.ac.uk/Lucene/Search.aspx (accessed 28 January 2011).

I asked for any studies with the keyword 'risk', a topic that was attracting a lot of attention at the time I was revising this chapter, in the title (see Plate 14.1). I clicked on the **Search** button. This resulted in nine studies being found (see Plate 14.2). This allows you to examine the documentation relating to any promising candidates for analysis and even to order the data set concerned. The

information provided typically gives a description of the study, along with a variety of particulars: sponsors; sampling details; method of data collection; main topics of the survey; and information about publications deriving from the study. It also informs you whether there are special conditions relating to access. With the one I specified, which was SN 5357—Public Risk Perceptions, Climate Change, and the Reframing of UK Energy Policy in Britain, 2005, I was told: 'The depositor has specified that registration is required and standard conditions of use apply.' Where there are special conditions, they normally involve signing an undertaking form. To order

Plate 14.2

Results of a search

Found 9 records Group by Series

	SN 4707 UEA-MORI Risk Survey, 2002 [ESDS Access and Preservation] This study investigated public attitudes towards science, risk and forms of governance. The centrepiece was a major quantitative survey of attitudes towards, and use of, alcohol, tobacco and other drugs by 15-16 year olds in Leicestershire with specific Study Description/Documentation Download/Order 
	SN 2738 Substance Abuse And Perceptions Of Risk : Young People's Attitudes To Personal Health, 1990 [ESDS Qualidata] Survey of attitudes towards, and use of, alcohol, tobacco and other drugs by 15-16 year olds in Leicestershire with specific Study Description/Documentation Download/Order 
	SN 5214 Divided We Stand: Bridging Differential Understanding Of Environmental Risk: GIS-P Maps, 2004 [ESDS Access and Preservation] The research project from which this dataset was produced was designed to help bridge the divide in understanding of the poss..... Study Description/Documentation Download/Order 
	SN 5357 Public Risk Perceptions, Climate Change And The Reframing Of UK Energy Policy In Britain, 2005 [ESDS Access and Preservation] The project consisted of undertaking a comprehensive empirical survey of public opinion towards future energy options for the..... Study Description/Documentation Download/Order 
	SN 5481 Confronting Unemployment: Families' Management Of Risk In A Flexible Labour Market, 1996-1999 [ESDS Qualidata] The data collected for this project comprise focus groups and in-depth interviews held separately with partners from both wag..... Study Description/Documentation See Access Information
	SN 5564 Eurobarometer 64.1: Mobility, Food Risk, Smoking, AIDS Prevention, And Medical Errors, September - October, 2005 [ESDS International Micro] The Eurobarometer (EB) survey series is a unique programme of cross-national and cross-temporal comparative social science res..... Study Description/Documentation Download/Order  Access via ZACAT
	SN 5831 Doing Youth Justice: Analysing Risk And Need Assessments In Youth Justice Practice, 2004-2005 [ESDS Qualidata] The past decade has seen a rapid expansion of the numbers of young people drawn into the criminal justice system and custody..... Study Description/Documentation Download/Order 
	SN 6072 Public Awareness Of Flood Risk: The Role Of The Environment Agency Flood Map, 2006-2007 [ESDS Access and Preservation] Public Awareness of Flood Risk: the Role of the Environment Agency Flood Map, 2006-2007 aims to investigate how official' in..... Study Description/Documentation Download/Order 
	SN 6295 Managing Food Chain Risks, 2006-2007 [RELU Data Support Service] This is a mixed method data collection. The study is part of the Rural Economy and Land Use (RELU) programme. RELU-RISK is 

the data, you need to have set up an account, which will have resulted in your being given a user name and password that will allow you to download data. You will need to find out if there is an administrative charge for receiving the data, but it is likely that, if you are a student at or a member of staff in a UK institution of higher education, there will be no charge or perhaps a nominal one. Online documentation is available towards the foot of the web page for the chosen study. This is not visible in see Plate 14.3, which also shows some of the information about the project that can be found on the ESDS web page, because you have to scroll further down the page presented in Plate 14.3. It should also be noted that Study SN 5357 is the source of the survey component of a mixed

methods study (Bickerstaff et al. 2008) referred to in Chapter 27.

Information about searching for qualitative data for the purpose of conducting a secondary analysis can be found in Chapter 24. Qualitative data can be searched for through ESDS Qualidata, which is also part of ESDS.

A useful starting point for many if not most of the most popular data sets that can be accessed through the ESDS can be found at:

www.esds.ac.uk/findingData/majorStudies.asp

(accessed 28 January 2011).

Table 14.1 lists several large data sets that are accessible to students and would repay further investigation in

Plate 14.3

Description and documentation for the chosen data

Abstract:

The project consisted of undertaking a comprehensive empirical survey of public opinion towards future energy options for the UK, with a particular emphasis on attitudes towards nuclear power when placed in the context of climate change.

The survey questionnaire consisted of 4 main sections. The first main section looked at climate change and nuclear power from a broad perspective, comparing these two with a range of other environmental and energy-related issues at different global/local scales. This section also examined attitudes towards various options for generating electricity. The second section specifically considered attitudes towards nuclear power. The third section examined attitudes towards climate change in more detail. Sections two and three contain a number of standardised questions that were aimed to measure general attitudes towards nuclear power and climate change, the perceived risks and benefits of the two issues, as well as questions on ambivalence, attitudinal certainty, and trust in risk regulation. In addition, the two sections contain a number of issue-specific questions. The fourth and final section of the questionnaire looked specifically at attitudes towards the reframing of nuclear power as a solution to climate change. This section contains questions that were designed to compare the risks of climate change with the risks of nuclear power, and attitudes towards different energy futures and options of electricity generation that might help to prevent climate change, using a split-sample technique.

Although there are a range of one-off or tracker surveys and risk perception studies available that have asked about energy generation, nuclear energy, and climate change separately, the value of the current instrument is that it generates a database which allows responses to be compared across all three sets of issues. In addition there is no comparable existing survey which asks, in a comprehensive way, about the reframing question.

Main Topics:

The main topics covered were: energy futures; nuclear power; climate change; environmental issues; risk perception.

Coverage:

Dates of Fieldwork: 01 October 2005–06 November 2005

Country: Great Britain

Spatial Units: Government Office Regions (GORs)

Observation Units: Individuals

Kind of Data: Numeric data

Universe Sampled:

Location of Units of Observation: National

Population: Members of the general public, aged 15 years and over in Great Britain during 2005

Methodology:

Time Dimensions: Cross-sectional (one-time) study

Sampling Procedures: Quota sample; Multi-stage stratified random sample

Number of Units: 1,491

Method of Data Collection: Face-to-face interview

Weighting: Data are weighted to age, sex, working status using 2001 census data (Variable= @WEIGHT0)

Language(s) of Written Materials:

Study Description: English

Study Documentation: English

Access:

Access Conditions: The depositor has specified that registration is required and standard conditions of use apply. The depositor may be informed about usage. See [terms and conditions](#) for further information.

Availability: ESDS Access and Preservation, UK Data Archive

Contact: Help desk: help@esds.ac.uk

Date of Release:

First Edition: 19 May 2006

Copyright:

Table 14.1

Large data sets suitable for secondary analysis

Title	Data set details	Topics covered
British Crime Survey (BCS)	Irregular survey of randomly selected sample of people questioned by structured interview. The 2001–2 BCS had a target sample of 36,000 individuals. Began in 1982 and since then carried out in 1984, 1988, and 1992. Biennially between 1992 and 2000 and annually from 2001.	Experience of and attitudes to crime; fear of crime; risk of crime.
British Household Panel Survey (BHPS)	A panel study that began in 1991 and is conducted annually by interview and questionnaire with a national representative sample of around 10,000 individuals in just over 5,000 households. The same individuals are interviewed each year. The BHPS is being replaced by the Understanding Society survey, which is based on a much larger panel of 40,000 households. See www.understandingsociety.org.uk (accessed 17 January 2011).	Household organization; labour market behaviour; income and wealth; housing; health; socio-economic values.
British Social Attitudes (BSA) Survey	More or less annual survey since 1983 of a representative sample aged 18 and over by interview and questionnaire. Each survey comprises an hour-long interview and a self-completion questionnaire.	Wide range of areas of social attitudes and behaviour. Some areas are core ones asked annually; others are asked irregularly.
Living Costs and Food Survey (LCFS)	The new name for the Expenditure and Food Survey, which began in April 2001 and which combined and replaced the FES and NFS. Data are based on households, which are asked to complete diaries of expenditure and income over a two-week period. In addition, face-to-face interviews are conducted by CAPI.	Areas formerly covered by FES and NFS.
Family Expenditure Survey (FES)	Annual interviews from 1957 with members of around 10,000 households who kept diary records of expenditure and income over two-week period. It was replaced by the Expenditure and Food Survey in 2001.	Family expenditure and income.
General Lifestyle Survey (GLF)	Formerly known as the General Household Survey (GHS), based on annual interviews since 1971 with members aged over 16 in over 8,000 randomly sampled households. No surveys were carried out in 1997–8, when the GHS was reviewed, and in 1999–2000, when it was revamped. Following this hiatus, the GHS has been resumed on an annual basis.	Standard issues such as education and health asked each year, plus additional items that vary annually. Huge variety of questions relating to social behaviour and attitudes.
Integrated Household Survey (IHS)	Also known as the Continuous Population Survey, this survey began in 2009. It comprises a set of core questions and modules from other continuous surveys carried out by the Office for National Statistics (ONS): Labour Force Survey, General Lifestyle Survey, Living Costs and Food Survey, English Housing Survey, and Life Opportunities Survey. For a short period, it also included the ONS Opinions Survey, but this module has now been dropped.	See: Labour Force Survey, General Lifestyle Survey, Living Costs and Food Survey, and ONS Opinions Survey.
Labour Force Survey (LFS)	Biennial interviews, 1973–83, and annual interviews, 1984–91, comprising a quarterly survey of around 15,000 addresses per quarter and an additional survey in March–May; since 1991 quarterly survey of around 60,000 addresses. Since 1998 core questions are also administered in member states of the European Union.	Hours worked; job search methods; training; personal details such as nationality and gender.
Millennium Cohort Study	Study of 19,000 babies and their families born between 1 September 2000 and 31 August 2001 in England and Wales, and between 22 November 2000 and 11 January 2002 in Scotland and Northern Ireland. Data were collected by interview with parents when babies were 9 months and around 3 years old. Since then, surveys have been conducted at ages 5 and 7 years old.	Continuity and change in each child's family and its parenting environment; assesses important aspects of the child's development.
National Child Development Study (NCDS)	Irregular but ongoing study of all 17,000 children born in Great Britain in the week of 3–9 March 1958. Since 1981 comprises interview and questionnaire. There have been six waves of data collection: 1965 (when members were aged 7 years), in 1969 (age 11), in 1974 (age 16), in 1981 (age 23), in 1991 (age 33), in 1999–2000 (age 41–2), in 2004 (age 46), and in 2008–9 (age 50–1).	Physical and mental health; family; parenting; occupation and income; housing and environment; etc.
National Food Survey (NFS)	First set up in 1940, this survey entailed interviews with a representative sample of households, and diary records for one week. In 1998 data were collected from both sources from 6,000 households. It was replaced by the Expenditure and Food Survey in 2001.	Nature of household; food shopping; meals.

Table 14.1

Continued

Title	Data set details	Topics covered
ONS Opinions Survey	Formerly, the ONS Omnibus Survey, a regular commercial survey by structured interview since 1990 of just under 2,000 people. Since 1990 it has been conducted eight times per year. It became the ONS Opinion Survey in 2008 when it was merged into the Integrated Household Survey (IHS).	Core questions each year about respondents, plus modules (asked on behalf of participating organizations) on topics that change annually concerning, e.g., food safety, eating behaviour, personal finance, sports participation.
Workplace Employment Relations Survey (WERS)	This survey has been carried out in 1980, 1984, 1990, 1998, and 2004. Workplaces of ten or more employees are sampled, and interviews carried out with managers, worker representatives, and employees. A new survey was conducted in 2011.	Pay determination; recruitment and training; equal opportunities; workplace change; work attitudes; management organization; employee representation; etc.

terms of their potential use in the context of research questions in which you might be interested. Further information about these data sets can be found via the ESDS or by using Google or any other search engine to look for them.

In discussing secondary analysis in this chapter, I have tended to emphasize large data sets like the BHPS. However, it is worth bearing in mind that the Data Archive holds data deriving from a wide range of studies. If you enter 'Research methods' as your search term when searching the catalogue, one of the studies that comes up is a content analysis of journal articles that

combine quantitative and qualitative research (SN 5077). This is a study that I conducted in 2003–4 and that forms the basis for much of Chapter 27. However, this is in no sense a large data set like the ones that have been the focus of attention for much of the previous discussion. It is also worth keeping a look-out for interesting good-quality studies that may not be archived. As mentioned in Chapter 13, the data from the Workplace Ethnography Project (Research in focus 13.4) can be easily downloaded and could form the basis of a very interesting secondary analysis project for someone with interests in areas like the sociology of work.



Official statistics

The use and analysis of official statistics for purposes of social research has been a very controversial area for many years. Agencies of the state, in the course of their business, are required to keep a running record of their areas of activity. When these records are aggregated, they form the official statistics in an area of activity. Thus, in Great Britain, the police compile data that form the crime rate (also known as 'notifiable crimes recorded by the police') and the Employment Service collects data that form the basis for the level of unemployment (also known as the 'claimant count'). These are just two, as it happens high-profile, sets of statistics that can be subsumed under the general category of 'official statistics'. Such statistics are frequently the cause of headlines in the mass media—for example, if there has been a sharp

increase in the level of recorded crime or unemployment. But they would also seem to offer considerable potential for social researchers.

We could imagine such official statistics offering the social researcher certain advantages over some other forms of quantitative data, such as data based on surveys.

- The data have already been collected. Therefore, as with other kinds of secondary analysis of data (see above), considerable time and expense may be saved. Also, the data may not be based on samples, so that a complete picture can be obtained.
- Since the people who are the source of the data are not being asked questions that are part of a research project, the problem of reactivity will be much less

pronounced than when data are collected by interview or questionnaire.

- There is the prospect of analysing the data both cross-sectionally and longitudinally. When analysing the data cross-sectionally, we could examine crime rates (and indeed the incidence of specific crimes) in terms of such standard variables as social class, income, ethnicity, age, gender, and region. Such analyses allow us to search for the factors that are associated with crime or unemployment. Also, we can analyse the data over time. Precisely because the data are compiled over many years, it is possible to chart trends over time and perhaps to relate these to wider social changes.
- There is the prospect as well of cross-cultural analysis, since the official statistics from different nation states can be compared for a specific area of activity. After

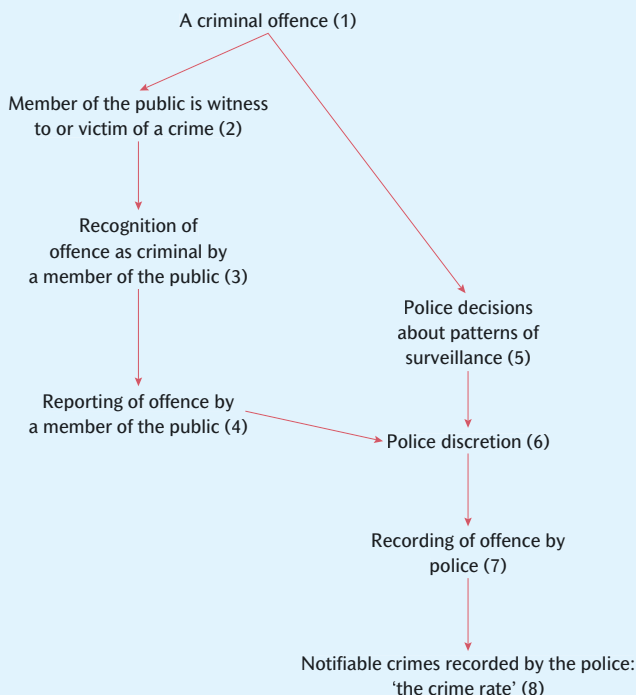
all, a sociological classic of the stature of Durkheim's *Suicide* (Durkheim 1952) was the result of a comparative analysis of official statistics on suicide in several countries.

However, readers who recall Thinking deeply 7.1 will already be on their guard. The official statistics concerned with an area of social life like crime can be very misleading, because they record only those individuals who are processed by the agencies that have the responsibility for compiling the statistics. Crime and other forms of deviance have been a particular focus of attention and concern among critics of the use of official statistics. Figure 14.1 illustrates, in connection with crime and the crime rate, the kinds of factor that can lead to concern.

If we take a criminal offence as the starting point (step 1), we can consider the factors that might or might

Figure 14.1

The social construction of crime statistics: eight steps



Source: adapted from a figure in Beardsworth et al. (n.d.)

not result in its becoming part of the crime rate. An offence might become a candidate for inclusion in the crime rate as a result of either of two events (it might be that others can be envisaged but these two represent major possibilities). First, the crime may be seen by a member of the public or a member of the public may be a victim of a crime (step 2). However, a crime has to be recognized as such before it will be reported to the police (step 3). Even if it is recognized as a criminal offence, the member of the public (even if he or she is a victim) may choose not to bring the crime to the notice of the police. This means that, if a criminal act goes unnoticed, or is noticed but not recognized as criminal, or is noticed and recognized as criminal but not reported to the police, it will not enter the official statistics. Step 4 is the reporting of the crime to the police. Even then the crime may not be entered into the crime statistics, because the police have considerable discretion about whether to proceed with a conviction and may choose to let the person off with a warning (step 6). They may be influenced by such factors as the severity of the crime, the perpetrator's previous record, the perpetrator's demeanour or suggestions of contrition, or their volume of work at the time.

Alternatively, a crime may be observed by the police as a result of their patterns of surveillance, which is a product of decisions about how best to deploy police officers (step 5). Once again, the crime may not become part of the crime rate because of the operation of police discretion. Thereafter, once the police exercise discretion in such a way as to lead them to seek a prosecution (step 6), the offence is recorded (step 7) and it becomes a 'notifiable crime recorded by the police' and as such part of the crime rate (step 8).

What are the implications of this process for the crime rate and for criminal statistics generally? For one thing, it means that a substantial amount of crime undoubtedly goes unrecorded as a result of: not coming to the attention of members of the public; not being recognized as criminal; not being reported; decisions regarding police surveillance that may result in some crimes being given lower priority; and the operation of police discretion. This undercurrent of unrecorded crime is frequently referred to as 'the dark figure' (Coleman and Moynihan 1996). Nor can crime be regarded as alone in this connection. Suicide statistics almost certainly fail to record many potential cases for inclusion and may even include a small number that are not in fact suicides (as a result of problems of deciding whether the 'victim' was involved in an accident or intended to commit suicide). It is extremely difficult to determine whether someone is the victim of suicide when there is no suicide note. Moreover, those responsible for concluding whether a death is a

suicide or not may come under considerable pressure not to record it as such, possibly because of feelings of potential stigma or because of religious taboos concerning suicide. To push the point even further, the deficiencies of official statistics do not relate just to areas of deviant behaviour such as crime and suicide. For example, the 'claimant count', which is used to gain a picture each month of the level of unemployment, may misrepresent the 'real' level of unemployment: people who are unemployed but who do not claim benefits or whose claim is disallowed will not be counted in the statistics, while those who form part of the claimant count but who work in part of what is known as the 'black' or 'informal' economy (and who therefore are not really unemployed) *will* be included in the unemployment statistics.

Increasingly, official statistics and summaries of them are available via the Internet. The Office for National Statistics is a good springboard for access to a wide variety of official statistics and can be accessed at: www.statistics.gov.uk/default.asp (accessed 28 January 2011).

Reliability and validity

Issues of reliability and validity seem to loom large in these considerations. Reliability seems to be jeopardized because definitions and policies regarding the phenomena to be counted vary over time. For example, the UK Government Home Office or police service policies may mean that more resources are to be put into surveillance of a certain area of crime, such as drugs or drink-driving. Moreover, as part of a crackdown, it may be that police officers are less likely to operate their discretion in such a way as to result in perpetrators being let off with a warning. The problem for the reliability of the crime statistics is that variations over time in levels of a particular crime may be due not to variations in the level of transgression but to variations in the propensity to expend resources on surveillance and to proceed with prosecution. Also, there may be changes over time in the definitions of crime or in the propensity of victims to proceed with a complaint. Such changes will clearly affect the degree to which fluctuations in the rate of occurrence of certain crimes reflect 'real' fluctuations in the incidence of those crimes or other factors (that is, variations over time in surveillance, operation of discretion, definition, propensity to report, and other factors). A further factor that can impair the reliability of crime statistics is 'fiddling' by police officers (see Thinking deeply 7.1 and 14.1). To the extent that such factors operate, the reliability of the crime data will be adversely affected, and, as a result, validity will be similarly impaired.



Thinking deeply 14.1

Fiddling the crime figures

An article in *The Times* (Leake 1998) reported that there was growing evidence that senior police officers frequently massage the crime statistics for their forces. The author argued that many officers deliberately 'lose' crimes in order to make their detection rates look better. As a result, crime rates are often lower than they should be. The article cites the following methods of suppressing crimes:

- classifying multiple burglaries—for example, in a block of flats in one day—as a single incident;
- cataloguing multiple credit card or cheque card frauds as a single offence;
- excluding common assaults, when people are not seriously injured, from the figures for violent crime;
- excluding drug offences where people admit to offending but are only cautioned.

These methods of reducing the crime rate will adversely affect reliability, because it is not possible to compare the figures over different time periods because of variations over time in the propensity to massage the data. Validity will be similarly affected, in part because measurement validity presupposes reliability (see Chapter 7) and also because the figures cannot be regarded as providing a picture of the true level of crime.

Also, the problems with official statistics extend to the examination of the variables with which the crime rate is associated. For example, it might be assumed that, if an examination of regional differences demonstrates that the crime rate varies by the chief ethnic or social class composition of an area, this implies that ethnic status and social class are themselves related to crime. There are two problems with drawing such an inference. First, there is an analytic difficulty known as the **ecological fallacy** (see Key concept 14.2). Second, even if we could ignore the problem of the ecological fallacy (which we cannot, of course), we would still be faced with an issue that is related to the matter of validity. Variations between ethnic groups or social classes may be a product of factors other than variations between ethnic groups and social classes in their propensity to commit crimes. Instead, the variations may be due to such factors as:

variations in the likelihood of members of the public reporting a crime when the perpetrator is of one ethnic group or class rather than others; variations in the surveillance activities of the police so that areas with a high concentration of members of one ethnic group or class rather than others are more likely to be the focus of activity; variations between ethnic groups or social classes in the propensity of police officers to exercise discretion; and problems for the police of learning about and investigating certain crimes that are themselves related to ethnicity or class (for example, white-collar crime). Similarly, as Douglas (1967) observes in connection with suicide statistics, it is quite likely that variations between ethnic and religious groups in suicide rates may be a product at least in part of variations in the predilection of the families of suicide victims to put pressure on official figures like coroners not to treat a death as a suicide.



Key concept 14.2

What is the ecological fallacy?

The ecological fallacy is the error of assuming that inferences about individuals can be made from findings relating to aggregate data. Coleman and Moynihan (1996) provide the example of the relationship between ethnicity and crime. They observe that findings showing a higher incidence of crime in regions with high concentrations of ethnic minorities have been used to imply that members of such minority groups are more likely to commit crimes. However, research on this issue in the 1960s suggested that in fact members of ethnic minority groups were less likely to offend. The fallacy can arise for several reasons—for example, it may not be the members of the minority groups who are responsible for the high levels of offending.

Condemning and resurrecting official statistics

In the 1960s, in particular, there was a torrent of criticism of various kinds of official statistics, especially those connected with crime and deviance. The various criticisms previously outlined were very much to the fore. In fact, so entrenched was the belief in many quarters that official statistics were of dubious value to social researchers that the view took root that they were virtually worthless. Instead, it was recommended that social researchers should turn their attention to the investigation of the organizational processes that produce the various deficiencies identified by the various writers. In the words of the writers of one influential article, rates of crime and other forms of deviance 'can be viewed as indices of organizational processes rather than as indices of certain forms of behavior' (Kitsuse and Cicourel 1963: 137). The effect of this view was to consign official statistics to the sidelines of social research so that they became an object of research interest rather than a potential source of data, although research based on official statistics continued in certain quarters. It would also be wrong to think that the critique formulated by writers such as Kitsuse and Cicourel (1963) and Douglas (1967) was the sole reason for the neglect of official statistics during this period. The fact that official statistics, because they are a sideline for many state agencies, are invariably not tailored to the needs of social researchers can be considered a further limitation. In other words, it may be that the definitions of apparently similar or identical terms (such as unemployment or socio-economic class) employed by those responsible for compiling official statistics may not be commensurate with the definitions employed by social researchers.

An important article by Bulmer (1980) questioned the relative neglect of official statistics by British sociologists in particular and represented a turning point in the views of many researchers towards this source of data (Levitas and Guy 1996). For one thing, Bulmer argued that the critique of official statistics had largely revolved around the elaboration of criticisms surrounding statistics relating to crime and deviance. He observes that these are subject to special well-known problems and it would be wrong to generalize these problems to the full range of official statistics. Many official statistics may be flawed in certain respects, but the flaws are not necessarily as pronounced as those to do with crime and deviance. Moreover, the flaws in many of the official statistics not concerned with crime and deviance are probably no worse than the errors that occur in much measurement deriving from methods like surveys based on

questionnaires and structured interviews. Indeed, some forms of official statistics are probably very accurate by almost any set of criteria, such as statistics relating to births, marriages, and deaths.

Bulmer also argues that, so far as some of the key variables in social research are concerned, the distance between the definitions employed by the compilers of official statistics and those employed by social researchers is not as great as is sometimes supposed. However, he notes that the case of social class is somewhat different. The development of the Registrar General's classification of social class groupings seems to have taken little notice of the schemes devised by sociologists, such as the influential Hall-Jones and Hope-Goldthorpe approaches. However, Bulmer notes that the Registrar General's classification nonetheless helpfully brings out clear divergences of socio-economic position between the groupings it comprises and is frequently employed by social researchers to make precisely this point. Bulmer points to the fact that data deriving from official statistics that show pronounced social class differences in mortality are extremely important and of considerable significance to medical and other sociologists. The data are not without problems and detractors, but there is a considerable willingness to use the statistics. The same applies to Inland Revenue data based on estate duties that have been employed to examine wealth distribution.

A further criticism of the rejection of various forms of official statistics is that it seems to imply that quantitative data compiled by social researchers are somehow error free or at least superior. However, as we have seen in previous chapters, while social researchers do their best to reduce the amount of error in their measurement of key concepts (such as through the standardization of the asking of questions and the recording of answers in survey research), it is not the case that the various measures that are derived are free of error. All social measurement is prone to error; what is crucial is taking steps to keep that error to a minimum. Therefore, to reject official statistics because they contain errors is misleading if in fact all measurement in social research contains errors. The problem here is that some official statistics are particularly prone to error, such as those relating to crime and deviance.

However, even here some caution is necessary. While data deriving from the British Crime Survey (BCS) may be employed to show that only a proportion of all crimes are notified to the police (see Thinking deeply 7.1), it would be wrong to conclude that the survey is an error-free yardstick. Coleman and Moynihan (1996) point to several measurement errors that are likely to afflict the BCS. For example, there is evidence to suggest that the

BCS results in an overestimation of serious incidents through a process known as ‘forward telescoping’. This means that serious incidents that are outside the recall period of twelve months (the period about which respondents are questioned) are erroneously considered to have occurred during that period. In other words, people have a tendency to believe that serious crimes of which they have been victims, but that occurred more than twelve months previously, actually occurred during the recall period. Coleman and Moynihan (1996) also point to errors arising from factors such as concealment in the course of interviewing. For example, there is some evidence to suggest that women are more likely to report sexual offences and domestic violence to the police (step 4 in Figure 14.1) than to a survey interviewer. In other words, dismissing official statistics on crime on the basis of survey evidence of the kind generated by the BCS is not without problems because it is not free of error itself.

It is clear that, following Bulmer’s (1980) statement of the issue, the wholesale rejection of official statistics by many social researchers has been tempered. While there is widespread recognition and acknowledgement that problems remain with certain forms of official statistics (in particular those relating to crime and deviance), each set of statistics has to be evaluated for the purposes of social research on its own merits.

Official statistics as a form of unobtrusive method

One of the most compelling and frequently cited cases for the continued use of official statistics is that they can be considered a form of unobtrusive measure, although

nowadays many writers prefer to use the term **unobtrusive method** (Lee 2000). This term is derived from the notion of ‘unobtrusive measure’ coined by Webb et al. (1966). In a highly influential book, Webb et al. argued that social researchers are excessively reliant on measures of social phenomena deriving from methods of data collection that are prone to *reactivity* (see Research in focus 3.3 and Key concept 12.4, where this idea is introduced). This means that, whenever people know that they are participating in a study (which is invariably the case with methods of data collection such as structured interviewing, self-administered questionnaire, and structured observation), a component of their replies or behaviour is likely to be influenced by their knowledge that they are being investigated. In other words, their answers to questions or the behaviour they exhibit may be untypical.

Official statistics fit fairly squarely in the second of the four types of unobtrusive measures outlined in Key concept 14.3. As noted in the box, this second grouping covers a very wide range of sources of data, which includes statistics generated by organizations that are not agencies of the state. This is a useful reminder that potentially interesting statistical data are frequently compiled by a wide range of organizations. An interesting use of such data is described in Research in focus 14.3. However, social researchers do not make a great deal of use of such data, and it is not irrelevant that the author referred to in the research presented in Research in focus 14.3 is an economist. There may be greater potential for searching out and mining statistical data produced by organizations that are relatively independent of the state.



Key concept 14.3 What are unobtrusive methods?

An unobtrusive measure is ‘any method of observation that directly removes the observer from the set of interactions or events being studied’ (Denzin 1970). Webb et al. (1966) distinguished four main types.

1. *Physical traces.* These are the ‘signs left behind by a group’ and include such things as graffiti and trash.
2. *Archive materials.* This category includes statistics collected by governmental and non-governmental organizations, diaries, the mass media, and historical records.
3. *Simple observation.* This refers to ‘situations in which the observer has no control over the behavior or sign in question, and plays an unobserved, passive, and nonintrusive role in the research situation’ (Webb et al. 1966: 112).
4. *Contrived observation.* This is the same as simple observation, but the observer either actively varies the setting in some way (but without jeopardizing the unobtrusive quality of the observation) or employs hidden hardware to record observations, such as video cameras.

Official statistics would be subsumed under Category 2, as would content analysis of media content of the kind described in Chapter 13. However, a content analysis like that described in Research in focus 13.4 would not be considered an example of an unobtrusive measure, because the material being content analysed (workplace ethnographies) derives from studies in which the data were generated in an obtrusive fashion. Structured observation of the kind covered in Chapter 12 will typically not fall into Categories 3 and 4, because the observer is usually known to those being observed. The study by Rosenhan (1973) described in Research in focus 12.5 is an illustration of contrived observation, because the pseudo-patients were not known to be researchers and they actively varied the situation by their own behaviour. The Daniel (1968) study described in Table 12.1 is also an example of contrived observation, because the 'actors' were not known to be researchers and by applying for rented accommodation they were actively varying the situation.

It is important to realize that Webb et al. (1966) were not intending that unobtrusive methods should supplant conventional methods. Instead, they argued that the problem they were identifying was the almost exclusive reliance upon methods that were likely to be affected by reactivity. Webb et al. argued for greater **triangulation** (see Key concept 17.4) in social research, whereby conventional (reactive) and unobtrusive (non-reactive) methods would be employed in conjunction. For example, they wrote that they were providing an inventory of unobtrusive methods, 'because they demonstrate ways in which the investigator may shore up reactive infirmities of the interview and questionnaire' (Webb et al. 1966: 174). In more recent years, many writers have preferred the term 'unobtrusive method' to 'unobtrusive measure' (e.g. Lee 2000), perhaps because the latter term is too suggestive of a quantitative research approach alone.

It is worth noting that unobtrusive methods encapsulate at least two kinds of ways of thinking about the process of capturing data. First, many so-called unobtrusive methods are in fact *sources* of data, such as graffiti, diaries, media articles, and official statistics. Such sources require analysis in order to be rendered interesting to a social scientific audience. Second, while documents of various kinds warrant being called 'unobtrusive' sources of data, in the sense that they have not been produced at the behest of a researcher (and are therefore not *reactive*), that should not be taken to mean that they are unproblematic. They are invariably produced for a purpose (albeit not specifically for research purposes) with an end in mind. Third, it includes *methods* of data collection, such as simple and contrived observation. While the data generated by such methods of data collection also require analysis, the data have to be produced by the methods. The data are not simply out there awaiting analysis in the way in which diaries or newspaper articles are (although, of course, a great deal of detective work is often necessary to unearth such sources). This means that neither of the terms 'unobtrusive methods' or 'unobtrusive measures' captures the variety of forms terribly well.

Lee (2000) has developed a classification of unobtrusive methods that differs slightly from that of Webb et al. (1966). He distinguishes the following kinds of data.

1. *Found data*. This category corresponds more or less exactly to *physical traces*.
2. *Captured data*. This category comprises both *simple observation* and *contrived observation*.
3. *Retrieved data: running records*. Records concerned with births, marriages, and deaths are prominent cases of this kind of record, whereby records can be examined over quite long periods so that changes can be explored. He also includes in this category personal advertisements like marriage announcements and dating advertisements (see Research in focus 13.2 for an example of research using the latter), as are job advertisements.
4. *Retrieved data: personal and episodic records*. With this category, Lee has in mind three kinds of data: **personal documents** (letters, diaries, memoirs), visual images in the mass media (for example, newspaper photographs and picture postcards), and 'documents produced through "institutional discovery" procedures' (Lee 2000: 87) (for example, reports of inquiries into the factors that led to a disaster).
5. Lee also distinguishes *records produced through the Internet*, especially the various forms of computer-mediated communication, such as email and various kinds of message boards and chat rooms. In the years since Lee wrote, blogs (web logs) might be added to this list.

Many of these different kinds of data are encountered elsewhere in this book—for example, personal documents in Chapter 23 and computer-mediated communications in Chapters 23 and 28. Each of the different types that Webb et al. and Lee distinguish poses distinctive questions in terms of issues like the reliability of the evidence and the ethical issues involved.



Research in focus 14.3

Using unofficial statistics? The case of New York taxi cab drivers

Following his informal observation on the behaviour of New York taxi drivers (cabbies), Camerer (1997) was interested in testing two different theories about the relationship between the number of hours a cabby works and average hourly earnings. One theory—the law of supply—predicted that cabbies would want to work more when their average hourly earnings would be high (for example, during bad weather or on working days when more business people are around) rather than when they are low. The second theory—daily income targeting—suggests that cabbies set themselves an income target for the day and once that target is attained they stop work for the day. On good days (when hourly wages are higher) this theory simply means they will stop earlier. Camerer obtained taximeter readings from the New York Taxi and Limousine Commission. The data allowed 3,000 observations of cabbies' behaviour for 1988, 1990, and 1994. Tips are not recorded, so a guess had to be made about likely levels in this area. The data provided unequivocal support for the daily income targeting theory. However, further analysis revealed a difference between newer and more experienced drivers: the former behaved very much in line with income targeting theory; the more experienced drivers were much more varied and their overall levels of behaviour were closer to the law of supply theory, though not entirely in conformity with it. Overall, if cabbies obeyed the law of supply, their mean incomes would rise by around 15 per cent.



Key points

- Secondary analysis of existing data offers the prospect of being able to explore research questions of interest to you without having to go through the process of collecting the data yourself.
- Very often, secondary analysis offers the opportunity of being able to employ high-quality data sets that are based on large reasonably representative samples.
- Secondary analysis presents few disadvantages.
- The analysis of official statistics may be thought of as a special form of secondary analysis but one that is more controversial because of the unease about the reliability and validity of certain types of official data, especially those relating to crime and deviance.
- The problems associated with official data relating to crime and deviance should not be generalized to all official statistics. Many forms of official statistics are much less prone to the kinds of errors that are detectable in relation to crime and deviance data, but there remains the possible problem of divergences of definition between compilers of such data and social researchers.
- Official statistics represent a form of unobtrusive method and enjoy certain advantages (especially lack of reactivity) because of that.



Questions for review

- What is secondary analysis?

Other researchers' data

- Outline the main advantages and limitations of secondary analysis of other researchers' data.

- Examine recent issues of one of the British sociology journals, like *Sociology*. Locate an article that uses secondary analysis. How well do the advantages and limitations you outlined fit this article?
- Does the possibility of conducting a secondary analysis apply only to quantitative data produced by other researchers?

Official statistics

- Why have many social researchers been sceptical about the use of official statistics for research purposes?
- How justified is their scepticism?
- What reliability and validity issues do official statistics pose?
- What are unobtrusive methods or measures? What is the chief advantage of such methods?



Online Resource Centre

www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Visit the Online Resource Centre that accompanies this book to enrich your understanding of secondary analysis and official statistics. Consult web links, test yourself using multiple choice questions, and gain further guidance and inspiration from the Student Researcher's Toolkit.